

SYSTEM INTERFACE FOR CELL AND/OR PACKET TRANSFER
AT AGGREGATE DATA RATES OF UP TO 10 GB/s

FIELD

5 The present invention relates to a POS-PHY Level 4 interface for cell and/or packet transfer.

BACKGROUND

10 Telecommunications equipment hardware for transmission and switching of packetized data are typically realized by interconnections of various integrated circuits. One or more of these integrated circuit devices are used to implement the functions of a layer in the communications protocol stack. The physical layer and link layer functions are typically implemented in separate devices that communicate with each other over a common interface. Such an interface must be designed to enable reliable and efficient data transfer. Various interfaces have been developed over the past few years to meet requirements of particular applications at the time. Recently, some interface proposals have been made to address requirements of packet and ATM cell transfer at data rates of up to 10 Gb/s (OC-192 SONET/SDH and 10 Gb/s Ethernet). The first is "UTOPIA Level 4" af-phy-0144.001, ATM Forum (ref. 1) and the second is "Proposal for a Common System Physical Interface Level 4 (SPI-4) to Support Physical Line Rates of up to 10 Gbps" (also known as SPI-4 Phase 1) Oif99.127, Optical Internetworking Forum (ref. 2).

15

20

25

The following definitions will be useful in the discussions to follow. The transmit data path refers to the signals associated with data transfer from the Link Layer to the physical layer (PHY) device. Conversely, the receive data path
 5 refers to transfer from the PHY device to the Link Layer device.

In ref. 1, control and status signals are sent together with payload data (i.e., sent "in-band") in a 32-bit wide data path. As shown in Fig. 1, there are separate transmit and receive
 10 data paths. Data (tx_data[31:0], rx_data[31:0]), control (tx_ctrl, rx_ctrl), and clock (tx_clk, rx_clk) lines are all implemented using LVDS I/O, operating at a nominal frequency of 400 MHz. Control and status for transmit data transfer is sent in the receive data path. Likewise, control and status for receive
 15 data transfer is sent in the transmit data path. When a control line is high in any given cycle, the corresponding data lines contain control information. Conversely, when a control line is low, the data lines contain payload data. The format of the control information is shown in Fig. 2.

In ref. 2, control and status signals are sent on separate lines (i.e., sent "out-of-band") from the 64-bit data path as shown in Fig. 3. All signals are implemented with HSTL Class 1 I/O, operating at a nominal frequency of 200 MHz. Flow
 25 control is accomplished by indicating FIFO full/not full status in a round-robin manner as shown in Fig. 4. When high, TxStart

and RxStart indicate the start of a round-robin sequence. Other signal definitions are summarized as follows. TxClk and RxClk are the clock lines. TxAddr[n:1] and RxAddr[n:1] indicate the port associated with the data transfer. TxData[63:0] and RxData[63:0] are used to carry payload data. TxPrty[3:0] and RxPrty[3:0] are parity bits calculated over the TxData and RxData lines. TxSOCP and RxSOCP indicate the start of a cell or packet. TxEOP and RxEOP indicate the cycle containing an end of packet. TxSize[3:0] and RxSize[3:0] indicate which bytes of a 64-bit word are valid upon end of packet. TxValid and RxValid indicate when the corresponding data lines are valid. TxError and RxError are used to indicate the occurrence of an error condition upon end of packet.

In principle, both ref. 1 and ref. 2 could be used for data transfer in 10 Gb/s applications. However, they suffer from various limitations. Both schemes have very high pin counts (136 and 164 pins for ref. 1 and ref. 2, respectively), which together with pins required for other functions, require larger IC package sizes. Moreover, their high pin counts inherently limit the number of interfaces that can be included in a given device, compared to a lower pin count interface. Both ref. 1 and ref. 2 also consume a relatively high amount of power (estimated at 4.08 and 3.42 W respectively, not including power consumption at the receivers). Both schemes specify only binary (i.e., full / not full) FIFO status information; more detailed FIFO status

information could be used by the scheduler at the far end to arrange data transfers to various ports so as to optimize transfer efficiency and to better avoid FIFO overflow and underflow conditions.

5

Since FIFO status information is sent in-band with data in ref. 1, it is difficult to use such an interface with unidirectional devices on either end (e.g., separate transmit and receive devices on either the PHY or the Link Layer side of the interface). An example for the case of unidirectional link layer devices is shown in Fig. 5, where an interface between those devices is required for sending FIFO status. For example, flow control information from the PHY for the Transmit Link Layer device would be received by a Receive Link Layer device, which would then send the information onto the Transmit Link Layer device to complete the feedback loop. Flow control information from the Receive Link Layer device itself would also need to be sent to the Transmit Link Layer device for in-band transmission to the PHY device. This additional interface between the Transmit and Receive Link Layer devices requires not only additional pins (along with attendant additional power consumption), but will also require additional engineering effort to specify and validate.

25

The control signaling protocol in ref. 1 permits arbitrary insertion of control words at any point during data

transfer. Such a protocol is needlessly complicated for the requirements at hand and is difficult to verify. Moreover, since several control words can elapse between one that contains parity information, the control signal (tx_ctrl / rx_ctrl) itself must also be included in the parity calculation if it is also to be protected. It is much simpler to have control words inserted only between data transfers, and to have parity contained in each control word.

With ref. 2, the parity bits protect only the data lines. It would be beneficial to have error protection over not only the data but also the associated control and status lines.

Accordingly, it is an object of the present invention to provide data recovery in the presence of skew between parallel data lines. It is a further object to provide a system interface for packet and cell transfer for OC-192 SONET/SDH and 10 Gb/s Ethernet Applications that have a lower pin count. It is yet another object of the invention to provide such an interface which has a lower power consumption. Another object of the invention is to provide such an interface having a simplified operation with minimal overhead, and better signal integrity. Finally, it is a further object of the invention to provide a system interface which is suitable for unidirectional link layer device implementations with error protection on data and control signals and 3-state flow control indication.

SUMMARY OF THE INVENTION

According to the invention there is provided a method of interfacing for packet and cell transfer between a first layer device and a second layer device, which includes dividing control information into an in-band portion and an out-of-band portion, transmitting the in-band portion in the data path from one of the first and second layer devices to another of said first and second layer devices, and transmitting the out-of-band portion outside of the data path from the another of the first and second layer devices to one of the first and second layer devices.

The in-band portion may be transfer-specific information and the out-of-band information may be FIFO status flow control information so that the interface operates independently in both transmit and receive directions.

A "1 1" framing pattern may be used on a FIFO status channel to mark boundaries of the framing pattern without requiring an out-of-band framing signal.

A training control pattern may be sent sufficiently often in order to allow a receive interface to check and correct for de-skew on start-up and during regular operation to compensate for skew variations due to changes in voltage, temperature, noise and other factors.

The FIFO status channel clock in the direction opposite to the corresponding data path may be used as a reference source for the data path clock transmitting from a side of the interface opposite to the transmitting end of the FIFO status channel. Conversely, the data path clock may be used as a reference source for the FIFO status channel clock transmitting from a side of the interface opposite to the transmitting end of the data path.

A transmitting end of the data path may send data and control signals precisely aligned with respect to a source-synchronous clock as well as the training pattern once every MAX_T, where MAX_T is configurable on start-up. This gives an implementation the flexibility of trading off overhead bandwidth required for sending training patterns and how frequently the receiver can check and correct for de-skew.

Each control word may contain an error-detection code and one or more control words inserted between bounded transfer periods whereby performance of the code is not degraded by overly long transfers.

An end-of-packet event and error codes may be combined into a two-bit code to reduce the number of bits required.

To further improve upon the efficiency of the interface, transfer information referring to a previous transfer and to a next transfer may be contained in one control word.

5 BRIEF DESCRIPTION OF THE DRAWINGS

Further features and advantages will be apparent from the following detailed description, given by way of example, of a preferred embodiment taken in conjunction with the accompanying drawings, wherein:

10

Fig. 1 is schematic diagram of a Utopia Level 4 Interface;

15

Fig. 2 is a schematic diagram illustrating the Utopia Level 4 control word format;

20

Fig. 3 is a schematic diagram of the System Physical Interface (SPI) Level 4 Phase 1;

25

Fig. 4 illustrates the SPI-4 FIFO status indication;

Fig. 5 is a schematic diagram showing an Interface between Transmit and Receive Link Layer Devices;

Fig. 6 is a general block diagram of a POS-PHY Level 4 Interface;

Fig. 7 is a schematic diagram showing the mapping of packets and ATM cells onto a payload stream;

5 Fig. 8 is a block diagram of the POS-PHY Level 4 Interface showing interface signals;

Fig. 9 is an example of a 43-byte packet payload transfer data structure;

10

Fig. 10 is an example of a 52-byte ATM cell payload transfer data structure;

15 Fig. 11 is an example of a 54-byte ATM cell payload transfer data structure;

Fig. 12 is an extended data structure;

20 Fig. 13 is a schematic diagram showing the range over which DIP-4 parity bits are computed;

Fig. 14 is a schematic diagram showing an example of DIP-4 odd parity encoding;

25 Fig. 15 is a data path timing diagram;

Fig. 16 is a FIFO Status Channel timing diagram;

Fig. 17 is a DIP-4 functional diagram showing calculation of the final DIP code;

5

Fig. 18 is a block diagram of the receiver;

Fig. 19 is a block diagram of the de-skew circuit; and

10

Fig. 20 is a block diagram of a T-clock FIFO Status Channel configuration.

DETAILED DESCRIPTION WITH REFERENCE TO THE DRAWINGS

1. Introduction

POS-PHY Level 4 is an interface for packet and cell transfer between a physical layer (PHY) device and a link layer device, for aggregate bandwidths of OC-192 ATM and Packet over SONET/SDH (POS), as well as 10 Gb/s Ethernet applications. This section provides a general overview of the interface. The next section contains more detailed descriptions of the signals and associated operations, data structures, start-up and AC timing parameters.

25

The following is a general synopsis of the POS-PHY Level 4 interface. For reference, a general block diagram is

shown in Fig. 6. POS-PHY Level 4 is the system packet interface for data transfer between the link layer and the PHY device; it is designed to meet requirements of this particular application, although it may be used in other applications as well.

5 "Transmit" and "Receive" refer, respectively, to data flow and associated control/status information for the Link Layer to PHY, and the PHY to Link Layer directions.

10 On both the transmit and receive interfaces, FIFO status information is sent separately from the corresponding data path. By taking FIFO status information out-of-band, it is possible to decouple the transmit and receive interfaces so that each operates independently of the other. Such an arrangement makes POS-PHY L4 suitable not only for bidirectional but also for
15 unidirectional link layer devices.

In both the transmit and receive interfaces, the packet's address, delineation information and error control coding is sent in-band with the data.

20

POS-PHY L4 has the following general characteristics:

- Point-to-point connection (i.e., between single PHY and single Link Layer device).

25

- Support for 256 ports (suitable for STS-1 granularity in SONET/SDH applications (192 ports), and Fast Ethernet granularity in Ethernet applications (100 ports)).

- 5
- Transmit / Receive Data Path:
 - 16 bits wide.
 - In-band port address, start/end-of-packet indication, error-control code.
 - LVDS I/O (IEEE 1596.3 - 1996 [1], ANSI/TIA/EIA-644-1995
- 10
- 700 MHz nominal data rate.
 - Source-synchronous double-edge clocking, 350 MHz nominal.
- Transmit / Receive FIFO Status Interface:
 - 87.5 MHz nominal clock rate.
 - CMOS I/O (3.3V)
 - 2-bit parallel FIFO status indication.
 - In-band Start-of-FIFO Status signal.
 - Source-synchronous clocking.
- 15
- 20

Data is transferred in bursts whose maximum size must be a multiple of 16 bytes. Information associated with each transfer (port address, start/end-of-packet indication and error-control coding) is sent in 16-bit control words described later in this

document. Fig. 7 shows how ATM cells and variable-length packets map onto the data stream.

2. Interface Description

5

Section 2.1 contains signal definitions for the transmit and receive directions. Section 2.2 describes the signal operation along with the data structures for payload data and in-band and out-of-band control/status information. Section
10 2.3 describes start-up parameters.

2.1. Signals

5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2

Table 2.1: POS-PHY Level 4 Interface Signal Summary

Signal	Direction	Description
TDCLK	Link to PHY	Transmit Data Clock. Clock associated with TDAT and TCTL. Data and control lines are driven off the rising and falling edges of the clock.
TDAT [15:0]	Link to PHY	Transmit Data. Used to carry payload data and in-band control words from the Link Layer to the PHY device. The control word format is described in Section 2.2.
TCTL	Link to PHY	Transmit Control. TCTL is high when a control word is present on TDAT[15:0]. It is low otherwise.

Signal	Direction	Description
TSCLK	PHY to Link	Transmit Status Clock. Clock associated with TSTAT.
TSTAT [1:0]	PHY to Link	Transmit FIFO Status. Used to carry round-robin FIFO status information, along with associated error detection and framing.
RDCLK	PHY to Link	Receive Data Clock. Clock associated with RDAT and RCTL. Data and control lines are driven off the rising and falling edges of the clock.
RDAT [15:0]	PHY to Link	Receive Data. Carries payload data and in-band control from the PHY to the Link Layer device. The control word format is described in Section 2.2.

Signal	Direction	Description
RCTL	PHY to Link	Receive Control. RCTL is high when a control word is present on RDAT[15:0]. It is low otherwise.
RSCLK	Link to PHY	Receive Status Clock. Clock associated with RSTAT.
RSTAT [1:0]	Link to PHY	Receive FIFO Status. Used to carry round-robin FIFO status information, along with associated error detection and framing.

2.2. Interface Operation and Data Structures

2.2.1. Data Path

5

As shown in Fig. 7, payload data is transferred in bursts. Packets longer than the configured maximum segment size must be segmented into multiple bursts. Control words are

inserted only between burst transfers. Once a transfer has begun, data words are sent uninterrupted until end-of-packet or the burst transfer size is reached, whichever comes first. The interval between the end of a given transfer and the next payload control word (marking the start of another transfer) consists of zero or more idle control words or training patterns (Section 2.2.3).

The minimum and maximum supported packet lengths are determined by the application. For ease of implementation however, successive start-of-packets must occur not less than 8 cycles apart (the gap between shorter packets is filled with idle control words).

Payload data bytes are transferred over the interface in the same order as they would be transmitted or received on the line side. The order of bytes within a word is shown in Figs. 9 to 11. The most significant bits (MSBs) of the constituent bytes correspond to bits 15 and 7. On payload transfers that do not end on a 16-bit boundary, the unused byte (after the last valid byte) is set to all zeroes. Fig. 9 shows an example of the data structure for payload transfer of an arbitrary packet. Figs. 10 and 11 show the corresponding data structures for a 52-byte ATM cell (i.e., without HEC field) and a 54-byte cell (with HEC and UDF).

A common control word format is used in both the transmit and receive interfaces. Table 2.2 describes the fields in the control word. When inserted in the data path, the control word is aligned such that its MSB is sent on the MSB of the transmit or receive data lines. A payload control word that separates two adjacent burst transfers contains status information pertaining to the previous transfer and the following transfer.

In order to support addressing requirements beyond 256 ports, as well as additional control functions, an extended data structure is defined, as shown in Fig. 12, in which the first data word of the payload data structure contains a 16-bit extended field. Usage of this field, either by itself, or in conjunction with the 8-bit port address of the preceding control word, is application-specific and beyond the scope of this specification. The extended address field counts as part of the payload transfer size.

Table 2.2. Description of Fields in the Control Words

Bit Position	Label	Description
15	Type	<p>Control Word Type.</p> <p>Set to either of the following values:</p> <p>1: payload control word (payload transfer will immediately follow the control word).</p> <p>0: idle or training control word (otherwise).</p>
14:13	EOPS	<p>End-of-Packet (EOP) Status.</p> <p>Set to the following values below according to the status of the immediately preceding payload transfer.</p> <p>0 0: Not an EOP.</p> <p>0 1: EOP Abort (application-specific error condition).</p> <p>1 0: EOP Normal termination, 2 bytes valid.</p> <p>1 1: EOP Normal termination, 1 byte valid.</p> <p>EOPS is valid in the first control word following a burst transfer. It is ignored and set to "0 0" otherwise.</p>

Bit Position	Label	Description
12	SOP	<p>Start-of-Packet.</p> <p>Set to 1 if the payload transfer immediately following the control word corresponds to the start of a packet. Set to 0 otherwise.</p> <p>Set to 0 in all idle and training control words.</p>
11:4	ADR	<p>Port Address.</p> <p>8-bit port address of the payload data transfer immediately following the control word. None of the addresses are reserved (all are available for payload transfer).</p> <p>Set to all zeroes in all idle control words.</p> <p>Set to all ones in all training control words.</p>
3:0	DIP-4	<p>4-bit Diagonal Interleaved Parity.</p> <p>4-bit odd parity computed over the current control word and the immediately preceding data words (if any) following the last control</p>

Bit Position	Label	Description
		word.

Table 2.3. Some Examples of Control Words

Control Word	Description
0110 00000000 abcd	Idle control word following end of transfer. End-of-packet, normal termination, 1 byte valid in last data word. (Note: abcd bits depend on contents of this control word and preceding transfer.)
0000 00000000 1111	Idle control word preceded by another (idle) control word.
1101 00000101 abcd	Payload control word following end of transfer. End-of-packet, normal termination, 2 bytes valid in last data word of preceding transfer (abcd bits depend on contents of this control word and preceding transfer). Start-of-packet in next transfer to port 5.

Fig. 13 shows the range over which the DIP-4 parity bits are computed. Fig. 14 shows a sample DIP-4 calculation; a

more detailed functional description is given in Section 3. The
 DIP-4 code is almost as easy to implement as a conventional BIP
 code. In the presence of random errors, it offers the same error
 protection capability as a comparable BIP code, but has an
 5 additional advantage of spreading single-column errors (as might
 occur in a single defective line) across the parity bits.

A timing diagram of the data path signals is shown in
 Fig. 15. This diagram is applicable to either the transmit or
 10 the receive interface. TCTL/RCTL is high when TDAT/RDAT contain
 control words. Idle periods correspond to back-to-back control
 words.

The actual clock rate used in practice is determined by
 the application at hand.

2.2.2. FIFO Status Channel

FIFO status information is sent periodically over the
 20 TSTAT serial link from the PHY to the Link Layer device, and over
 the RSTAT serial link from the Link Layer to the PHY device.
 Implementation of the FIFO status channel for the transmit
 interface is mandatory; the corresponding implementation for the
 receive interface is optional. If both status channels are
 25 implemented, they shall operate independently of each other. The
 FIFO status of each port is encoded in a 2-bit data structure,

whose format is defined in Fig. 12. The most significant bit of each port status is sent over TSTAT / RSTAT [1], while the least significant bit is sent over TSTAT / RSTAT [0]. Section 2.3 discusses configuration of the FIFO status port sequence. The port sequences on the transmit and receive interfaces may be configured differently from each other. The "1 1" pattern is reserved for in-band framing; it must be sent once prior to the start of the FIFO status sequence. A DIP-2 odd parity checksum is sent at the end of each complete sequence, immediately before the "1 1" framing pattern. The DIP-2 code is computed over all preceding FIFO status indications sent after the last "1 1" framing pattern, diagonally over TSTAT / RSTAT [1] and TSTAT / RSTAT [0]. A functional description of the DIP-2 code is given in Section 3. While the parity bits can mimic the "1 1" pattern, the receiving end can still frame successfully by synchronizing onto the last cycle in a repeated "1 1" pattern and by making use of the configured length of the sequence (see Section 2.3) when searching for the framing pattern. A timing diagram of the FIFO status channel is shown in Fig. 16. To permit more efficient FIFO utilization, the MaxBurst1 and MaxBurst2 credits are granted and consumed in increments of 16-byte blocks. A burst transfer shorter than 16 bytes (e.g., end-of-packet fragment) will consume an entire 16-byte credit.

A continuous stream of repeated "1 1" framing patterns may be sent to indicate an error condition. For example, it may

be sent to indicate that data path de-skew (Section 2.2.3) has not yet been completed or confirmed. When a repeated "1 1" pattern is detected, all outstanding credits are cancelled and set to zero.

5

The indicated FIFO status is based on the latest available information. A STARVING indication provides additional feedback information, so that transfers can be scheduled accordingly. Applications which do not need to distinguish between HUNGRY and STARVING need only examine the most significant FIFO status bit.

10

2.2.3. Training Pattern for Data Path Deskew

A training pattern is sent at least once every preconfigured bounded interval on both the transmit and receive interfaces. These training patterns may be used by the receiving end of each interface for de-skewing bit arrival times on the data and control lines. The procedure defined in this section is designed to allow the receiving end to correct for relative skew differences of up to +/- 1 bit time. The training pattern consists of 1 idle control word, 10 (repeated) training control words and 10 (repeated) predefined data patterns. The idle control word removes dependencies of the DIP-4 in the training control words from preceding data words. Assuming a maximum of +/- 1 bit time in bit alignment jitter on each line, and a maximum of +/- 1 bit time relative skew between lines, there will

15

20

25

be at least 8 bit times during which a receiver can detect a training control word prior to de-skew. The predefined data pattern is chosen to be orthogonal to the training control word. In the absence of bit errors in the training pattern, a receiver should be able to successfully de-skew the data and control lines with one training pattern.

The sending side of the data path on both the transmit and receive interfaces must send the training pattern in Table 2.5 (in sequence from cycles 1 through 21) at least once every MAX_T payload or idle control words have been sent, where MAX_T is configurable on start-up. Training patterns at the transmit and receive interfaces are scheduled independently. They must not be inserted within a payload burst transfer (i.e., not inserted between a payload control word and any of the subsequent data words until the end of transfer).

Table 2.5. Training Pattern

(Note: In cycle 1, XX and abcd depend on the contents of the interval after the last preceding control word.)

Cycle	TCTL / RCTL	TDAT[i] / RDAT[i]															
		1	1	1	1	1	1	9	8	7	6	5	4	3	2	1	0
		5	4	3	2	1	0										
1	1	0	X	X	0	0	0	0	0	0	0	0	0	a	b	c	d
2-11	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
12-21	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0

2.3. Start-Up Parameters

The sequence of ports at a FIFO status channel is defined in a data structure called CALENDAR, where CALENDAR[i], i = 1,...,CALENDAR_LEN, refers to the ith port in the repeating sequence. In general, CALENDAR_LEN corresponds to the number of ports with the lowest data rate that can be accommodated in the total data rate of the given application. The calendar sequence (of length CALENDAR_LEN) is repeated CALENDAR_M times before the

DIP-2 parity and "1 1" framing words are inserted. CALENDAR_LEN and CALENDAR_M are both greater than zero.

Examples:

- 5 1. Single OC-192 or 10 Gb/s Ethernet port: CALENDAR_LEN = 1,
 CALENDAR[1] = 1.
2. Four OC-48 ports: CALENDAR_LEN = 4, CALENDAR[i] = 1, 2, 3,
 4.
3. Two OC-48 channels (ports 1 and 2), eight OC-12 channels
 10 (ports 3 through 10): CALENDAR_LEN = 16, CALENDAR[i] = 1, 2,
 3, 4, 1, 2, 5, 6, 1, 2, 7, 8, 1, 2, 9, 10,... Other
 combinations are feasible: CALENDAR[i] = 1, 3, 2, 4, 1, 5,
 2, 6, 1, 7, 2, 8, 1, 9, 2, 10,
4. Ten 1 Gb/s Ethernet ports: CALENDAR_LEN = 10, CALENDAR[i] =
 15 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

 The values of CALENDAR_LEN, CALENDAR_M and CALENDAR[i],
 i = 1,...,CALENDAR_LEN, must be identical in both the PHY and Link
 Layer devices for each interface. They need not be identical on
 20 both the transmit and receive FIFO status channels. The maximum
 supported value of CALENDAR_LEN is contained in the parameter
 MAX_CALENDAR_LEN, whose upper bound is implementation-specific.
 MAX_CALENDAR_LEN need not be identical on either side of the
 transmit or receive FIFO status channels. Users however, must
 25 ensure that the value of CALENDAR_LEN on the sending side of a

FIFO status channel must not exceed MAX_CALENDAR_LEN on the receiving side.

For the FIFO Status channel(s), MaxBurst1 and MaxBurst2
 5 may be configured to apply globally over all ports, or to apply on a per-port basis. In either case, both parameters must be consistently configured at the PHY and Link Layer devices for each interface, but need not be identical between the transmit and receive interfaces.

10 For the data path de-skew procedure, MAX_T is configured only on the sending side of the data paths on the transmit and receive interfaces. MAX_T need not be identical over both interfaces.

3. Functional Descriptions of the DIP-4 and DIP-2 Codes

This section gives a brief description on how the DIP-4 and DIP-2 codes can be implemented. As shown in Fig. 14, the parity bits are summed diagonally across the parallel lines such
 20 that the total number of 1's (counting parity) is an odd number.

For the DIP-2 code used in a FIFO status channel, the computation involves summing diagonally over two columns (TSTAT / RSTAT [1], TSTAT / RSTAT [0]) over all status words since the last "1 1" framing pattern. For the sake of consistency, the
 25 status lines relative to Fig. 14 are arranged with the MSB and

LSB as the left and right columns respectively, with time moving downward. The parity bits at the bottom are set such that the total number of ones in each diagonal sum (including parity) is an odd number.

5 For the DIP-4 code, assume that the stream of 16-bit data words are arranged as shown in Fig. 17, MSB at the leftmost column, time moving downward. As in Fig. 14, the parity bits are generated by summing diagonally (in the control word, the space occupied by the DIP-4 code is set to all 1's during encoding).

10 The first 16-bit checksum is split into two bytes, which are added to each other modulo-2 to produce an 8-bit checksum. The 8-bit checksum is then divided into two 4-bit nibbles, which are added to each other modulo-2 to produce the final DIP-4 code.

15 4. Functional Description of Deskew Operation at the Receiver

This section describes a method for implementing data path de-skew using the pattern defined in Table 2.5. The reader is referred to Figs. 18 and 19 for reference points.

20 In Fig. 18, each of the 17 lines in the data path (16 lines of data (DATA_IN[15:0]), 1 line for control (CTRL_IN)), are independently centered with respect to the received clock (CLK_IN) by data recovery unit (DRU) circuits. Each DRU outputs their respective sampled bits onto de-skew circuit at reference point A.

In Fig. 19, each of the 17 lines in reference point A are fed into serial-in parallel-out (SIPO) blocks (B1 through B17) that emit n-bit words at a rate of $1/n$ the rate of the serial inputs ($n=N$ in Fig. 19). The outputs of blocks B1 to B17 are then fed into a plurality of registers C1 through CN, where N is an integer. At each updated output of the SIPO blocks, registers C1 through CN contain (possibly skewed) samples of the 17-bit data path at time periods, t , $t+1$, $t+2$, and $t+N-1$. C0 contains the sample at time $t-1$ (which may be obtained from the contents of CN in the previous SIPO update). When an incoming training pattern (as defined in Table 2.5) is received, one or more of the registers C1 to CN should contain a training control word (due to the repetitions of the training control word, and assuming a sufficiently small bit error rate and skew). The contents of C1 through CN are fed into a training detector block (reference point D), which declares detection of a training pattern based on the contents of C1 to CN. The detection scheme may be based on sampling a threshold number of training control words in C1 to CN. Other schemes may also be feasible. Once the training pattern is detected, 17 transition detection blocks (E1 to E17) independently search for the first 1->0 (or 0->1, depending on the bit line) transition that occurs after the last repetition of the training control word. Block Ei searches for the transition on the i th bit position from registers C0 through CN. The registers corresponding to each transition identified by E1 to E17 are fed to an aligner block (reference point F), which

selects the appropriate register (from C1 to CN) from which to read each bit in order to present a de-skewed output in reference point G.

5 As an alternative implementation of the DIP-2 code used at the end of the FIFO status sequence, the parity bits on the FIFO status lines can be defined such that the "1 1" framing pattern is never mimicked. One possible mapping is to use a two-cycle parity mapping as follows:

10

0	0	→	0	0	0	0
0	1	→	0	1	0	1
1	0	→	1	0	1	0
1	1	→	1	0	0	1

15

Data and control/status paths can be wider (running at correspondingly lower clock rates) or narrower (running at higher clock rates). For example, a 4-bit, 3.5 GHz interface may be used to support similar aggregate bandwidth applications. To support higher bandwidths like OC-768 (40 Gb/s) for example, a 64-bit wide 700 MHz or a 16-bit 3.5 GHz data path may be used. Other combinations of widths and clock rates may also be used for these and other data rates.

20

25 Persons skilled in the art will recognize that there are other input/output technologies, data and/or control path widths and clock frequencies which can also be used to implement

POS-PHY L4. There are other error control coding schemes that can be used in place of the DIP-4 and DIP-2 codes. Framing for the FIFO status channels can also be implemented out-of-band by using a separate line to provide a frame pulse at the start of the calendar sequence. The transmit and/or receive interfaces can also be operated completely open-loop, without the use of the corresponding FIFO status channels. Clock lines on the data path and FIFO status channels can be eliminated by using a coding scheme like 8b10b (and/or scrambling) and performing associated clock recovery for each line at the receiver. The Link Layer device can exert flow control on the receive interface by means of a receive enable (RENB) signal. When RENB is high, the PHY may transfer as much data possible over the receive interface to the Link Layer device. When RENB is low, the PHY device can transfer only a predefined amount of data until RENB returns back to high. The STARVING indication in the FIFO status channel may be triggered by conditions which indicate that an underflow is imminent, such as the following:

(a) A packet is being transmitted over the line interface (i.e. the FIFO is being emptied);

(b) A corresponding EOP indication for that packet has not yet been received by the PHY device (i.e., not written into its FIFO);

(c) The FIFO level for the given port has dropped below a configured threshold (Starve_Threshold).

The data and control lines may be scrambled to break long runs of 1's or 0's and a T-clock configuration may be used on the FIFO status channel in place of source-synchronous clocking, as shown below in Fig. 20.

5

Accordingly, while this invention has been described with reference to illustrative embodiments, this description is not intended to be construed in a limiting sense. Various modifications of the illustrative embodiments, as well as other
10 embodiments of the invention, will be apparent to persons skilled in the art upon reference to this description. It is therefore contemplated that the appended claims will cover any such modifications or embodiments as fall within the true scope of the invention.

15